

- What are the long-term limitations in the scaling of grid middleware, services and interfaces?
- What are the long-term limitations in the scaling of authentication and identity management?
- How would we have to configure grid sites differently N years from now, in conditions where we might be operating many cores per node?
- How will experiments change their operations to carry out the same tasks with less personnel support?
- What are the long-term limitations in the scaling of current data management schemes?



Open Science Grid

Future of the Grid Security

Mine Altunay
OSG Security Officer
Fermilab National Laboratory





Move to the Certificate-Less Grid

- We all know the frustration and difficulty of using end user certificates.
- The question is ***how we move end users to a certificate-less grid.***
- Two projects going on that will ease the pain.
 - CILogon Basic CA. <https://cilogon.org/>
 - Still using certificates, but easier to manage than classic CAs
 - Online CA, works very similar to CERN CA.
 - No separate step to vet your identity. Integrated with end user's campus identity.
 - End user authenticates with their home organization in the normal way.



Open Science Grid

Move to the Certificate-Less Grid

- CILogon captures end user's identity info from the home institution.
- Happens under 1 minute
- We are promoting CILogon Basic CA in OSG.
- OSG central services (OIM, twiki, Docdb) already accepts CILogon Basic certs.
- We have 5 sites who agreed to accept these certificates. We are getting the word out to more sites and VOs
- If your VO or site can take advantage of CILogon Basic CA, please contact the security team
osg-security-team@opensciencegrid.org



Open Science Grid

Move to the Certificate-Less Grid



CILogon Service

Show
Help

Select An Identity Provider:

Emory University	▼
Florida International University	
Georgetown University	
Google	

Search:

Remember this selection: ☐

Log On

By selecting "Log On", you agree to [CILogon's privacy policy](#).



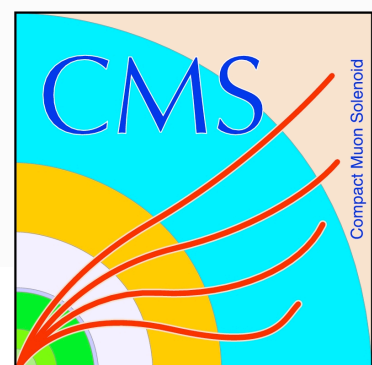
Move to the Certificate-Less Grid

- 2nd project: Running end user jobs without any user certificates
 - Allows VOs with job submission portals (glideinWMS, Panda, etc) to send user jobs without user certificates.
 - Traceability (whose job is this running on the WN) is an important question.
 - OSG created a traceability requirements document for such VOs (OSG-doc-1149 https://osg-docdb.opensciencegrid.org:440/cgi-bin/RetrieveFile?docid=1149;filename=OSG_Traceability_Requirements_v4.docx;version=1)
 - Encouraging Sites and VOs to give us feedback over the document:
 - VOs: can you meet these requirements?
 - Sites: can you accept these jobs?

Future of the Grid - Lightning Talks

OSG AHM 2013
13. March 2013

Oliver Gutsche



- ▶ How will experiments change their operations to carry out the same tasks with less personnel support?

- ▶ Currently the most manpower operation intensive tasks are:
 - ▶ Monitoring of systems
 - ▶ Solution of infrastructure problems (broken srm, hung transfer, corrupt file, stuck queue, black whole node, etc.)
 - ▶ Optimization of resource usage

- ▶ What we need: self-learning systems
 - ▶ Monitors the infrastructure and/or ongoing work
 - ▶ Learns resource requirements of work (memory, running time, local disk, bandwidth)
 - ▶ Optimizes the work to be scheduled according to what was learned (run at other sites, transfer file from different destination, ...)
 - ▶ Detects problems in the infrastructure, avoids scheduling new work, but schedules problem for repair
- ▶ What we need: self-healing systems
 - ▶ Either detect problems or receive external trigger for problem
 - ▶ Solves problems themselves (most cases) or flags the system for manual intervention
- ▶ Operations will become harvesting and analyzing information and helping the system to learn.

- ▶ What basic services that are anticipated to be needed for new experiments are missing right now?
- ▶ Growth in GRID computing will be in small non-HEP applications
 - ▶ Mostly relying on opportunistic resources
- ▶ We know a lot of opportunistic CPU resources
 - ▶ How to enable access.
 - ▶ How to use them.

- ▶ But we are lacking a clear picture for storage
- ▶ The very old statement "I can buy a disk at <insert favorite electronics store> and I have my storage" needs to become true, but in a GRID sense
- ▶ We need the OSG DropBox service
 - ▶ Every OSG user gets 10 GB free DropBox space, available on their desktop/laptop and available to all their GRID jobs
 - ▶ Users can buy more
 - ▶ Users can use the scientific storage appliance software kit: buy disk, install software, connect to network => available to all my GRID jobs and my desktops/laptops (risking loosing their data when their disks breaks of course, ...)

Future of the Grid Lightning Talk

Shawn McKee/University of Michigan

At the OSG AHM in Indianapolis, IN

March 13, 2013

Future of “the Grid”



- * **Quick review:** for me grids represent the infrastructure necessary to allow different “administrative” domains to share access to services, computers and storage with selected users or groups of users (virtual organizations)
- * How might this evolve? Might also ask:
 - ❑ Who needs it? Distributed (large) science domains (any group with distributed membership, computation/storage needs and distributed resources)
 - ❑ What is changing in technologies underpinning “grids”?
Virtualization, evolutionary improvements, revolutionary changes?
 - ❑ How is the pursuit of science changing? Bigger datasets, more interactions amongst scientists, more opportunities to explore alternate hypotheses, easier to focus on science instead of infrastructure.

Recent Changes: Clouds



- * I would claim that **Grids + Virtualization ~ Clouds**
- * **Clouds** are also primarily driven by commercial interests.
- * For grid users, clouds are not able to meet our primary needs:
 - ❑ Still not as cost-effective as our specially designed and tuned infrastructures that we share via “Grid” technologies
 - ❑ Especially problematic for I/O intensive workloads...no ways to easily control the relationship between storage and compute locations
 - ❑ They are able to provide CPU cycles suitable for some science workflows competitively (spot-market pricing)

The Future



- * Will some future variant of “Clouds” supplant “Grids”?...likely
- * However, I suspect we (science domain users) will still own and tune our (potentially) virtual resources.
- * The trend indicates these resources may become “rented” out of a larger pool rather than purchased and assembled at a local site or set of sites.
 - Those resources will initially be “vanilla” but will have the ability to learn about the characteristics of the workflows they run and self-tune/optimize their configuration in response, automatically
 - Even so, I believe there will be specialists among us who will need to understand options and provide overall guidance on managing and delivering the best infrastructure for specific Vos
- * **In physics, we will probably still call it “the Grid” ☺**

OSG Technology in 2018

- More heterogeneous resource allocation - via SSH, CE, and/or EC2.
 - We will co-allocate storage, CPU, RAM, and network resources. Number of allocated CPUs will not necessarily be equal to 1.
 - Batch systems will schedule/allocate 80% of resources, virtualized solutions will allocate 20%.
 - We will help users allocate resources across OSG and XSEDE.
- More homogeneous user environment. We will have improved methods for making the software and filesystems look similar for all worker nodes.
- Pilots will handle file staging for payloads. All non-LHC VOs will have files kept at one archival site and treat all other storage as cache.
 - Where is the archive? OSG? XSEDE? I don't know!
 - LHC VOs will evolve their data transfers, remote IO, and management to increasingly larger scales. I do not see a revolution or things breaking down!
- Users will know how to use HTCondor but not grid-proxy-init.

OSG in 2018

- We will be able to better answer “why is my job not running?”
- We will be able to deliver DHTC to “Joe the Biologist” without embarrassing ourselves.
 - We will need “Joe the Physicist” to find all the bugs first.
- We will run out of opportunistic resources. The day we realize this will be noisy and have a lot of shouting.
- OSG will deploy a centrally-managed fairshare or scheduling mechanism.

Grid Futures

Rob Gardner

Questions

- What are the long-term limitations in the scaling of grid middleware, services and interfaces?
- What are the long-term limitations in the scaling of authentication and identity management?
- How would we have to configure grid sites differently N years from now, in conditions where we might be operating many more cores per node?
- How will experiments change their operations to carry out the same tasks with less personnel support?
- What are the long-term limitations in the scaling of current data management schemes?
- What basic services that are anticipated to be needed for new experiments are missing right now?

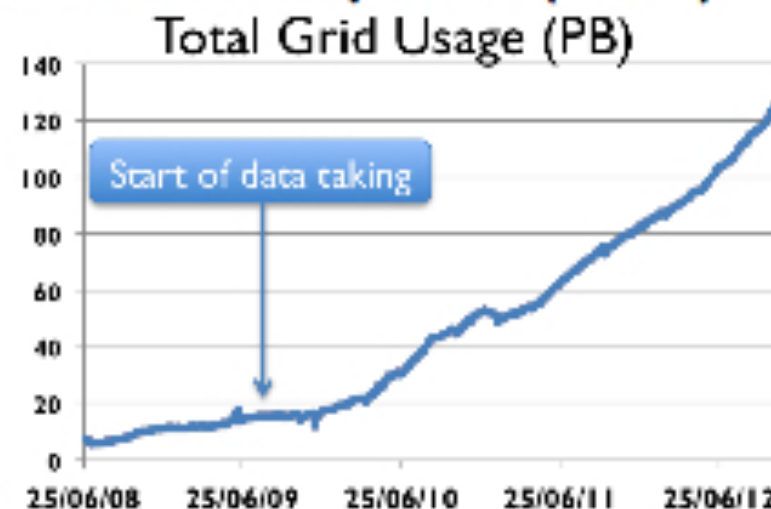
(Some) answers

- How will experiments change their operations to carry out the same tasks with less personnel support?
 - Make better systems
 - Take advantage of advances in the rest of the world

(Some) answers

- What are the long-term limitations in the scaling of current data management schemes?

- The Distributed Data Management project manages ATLAS data on the grid
- The current system is Don Quijote 2 (DQ2)
 - 130 Petabytes
 - 600k datasets
 - 355 million files
 - 800 active users
 - 130 sites
- DQ2 works, but ...
 - Scaling problems, heavy operational burden and difficulties to add new features and technologies



(Some) answers

- What basic services that are anticipated to be needed for new experiments are missing right now?
 - Software and data preservation
 - Sustained usability
 - Preserving processing environments
 - DASPOS, DPHEP, ..

